US 20160147705A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2016/0147705 A1

(43) Pub. Date: May 26, 2016

Arroyo et al.

(54) **IMPLEMENTING HEALTH CHECK FOR OPTICAL CABLE ATTACHED PCIE ENCLOSURE**

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION**, Armonk, NY (US)

(72) Inventors: **Jesse P. Arroyo**, Rochester, MN (US); **Ellen M. Bauman**, Rochester, MN (US); **Timothy R. Block**, Rochester, MN (US); **Christopher J. Engel**, Rochester, MN (US); **Kaveh Naderi**, Austin, TX (US); **Harald Pross**, Wildberg (DE); **Thomas R. Sand**, Rochester, MN (US)

(21) Appl. No.: **14/549,801**

(22) Filed: **Nov. 21, 2014**

**Publication Classification**

(51) **Int. Cl.**

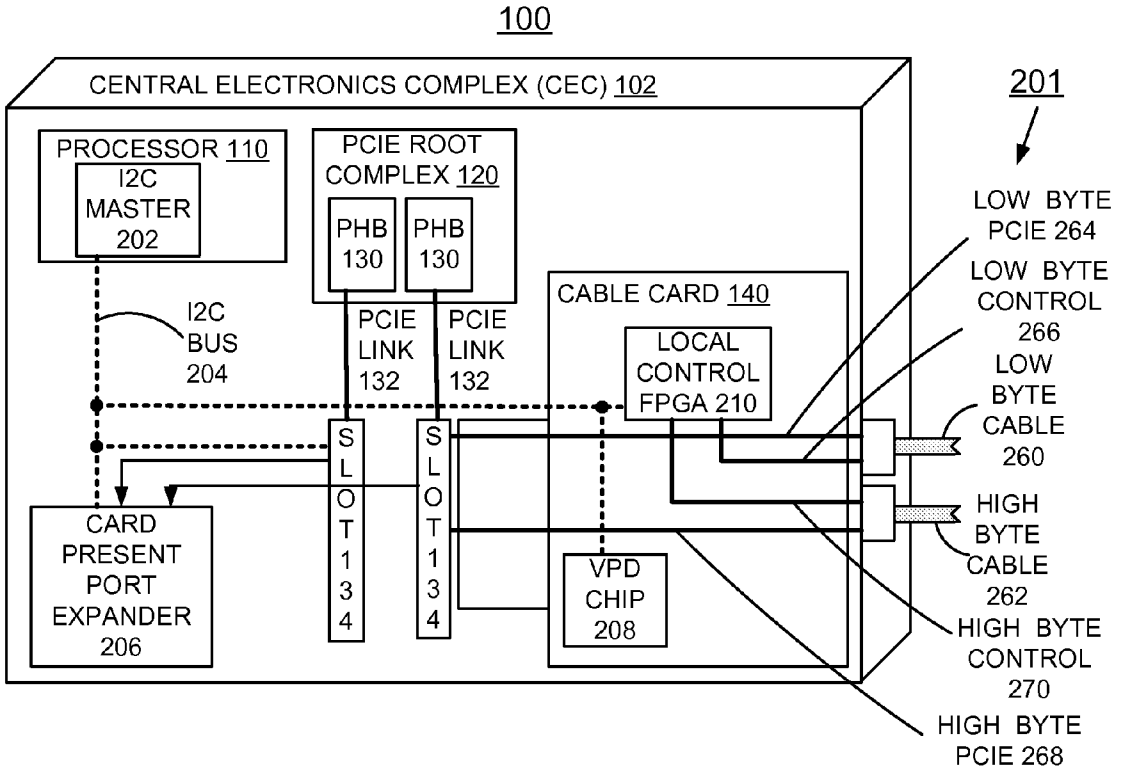
| | |
|---|---|
| ***G06F 13/42*** | (2006.01) |
| ***G06F 11/07*** | (2006.01) |
| ***G06F 9/445*** | (2006.01) |

(52) **U.S. Cl.**

CPC ..... ***G06F 13/4282*** (2013.01); ***G06F 9/44505*** (2013.01); ***G06F 11/0793*** (2013.01)

(57) **ABSTRACT**

A method, system and computer program product are provided for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCIE) enclosures in a computer system. System firmware is provided for implementing health check functions. One or more optical cables are connected between a host bridge and a PCIE enclosure. A PCIE link to the PCIE enclosure is reset responsive to a predefined event. After a set delay, a PCIE link health check is performed verifying PCIE link width and speed.



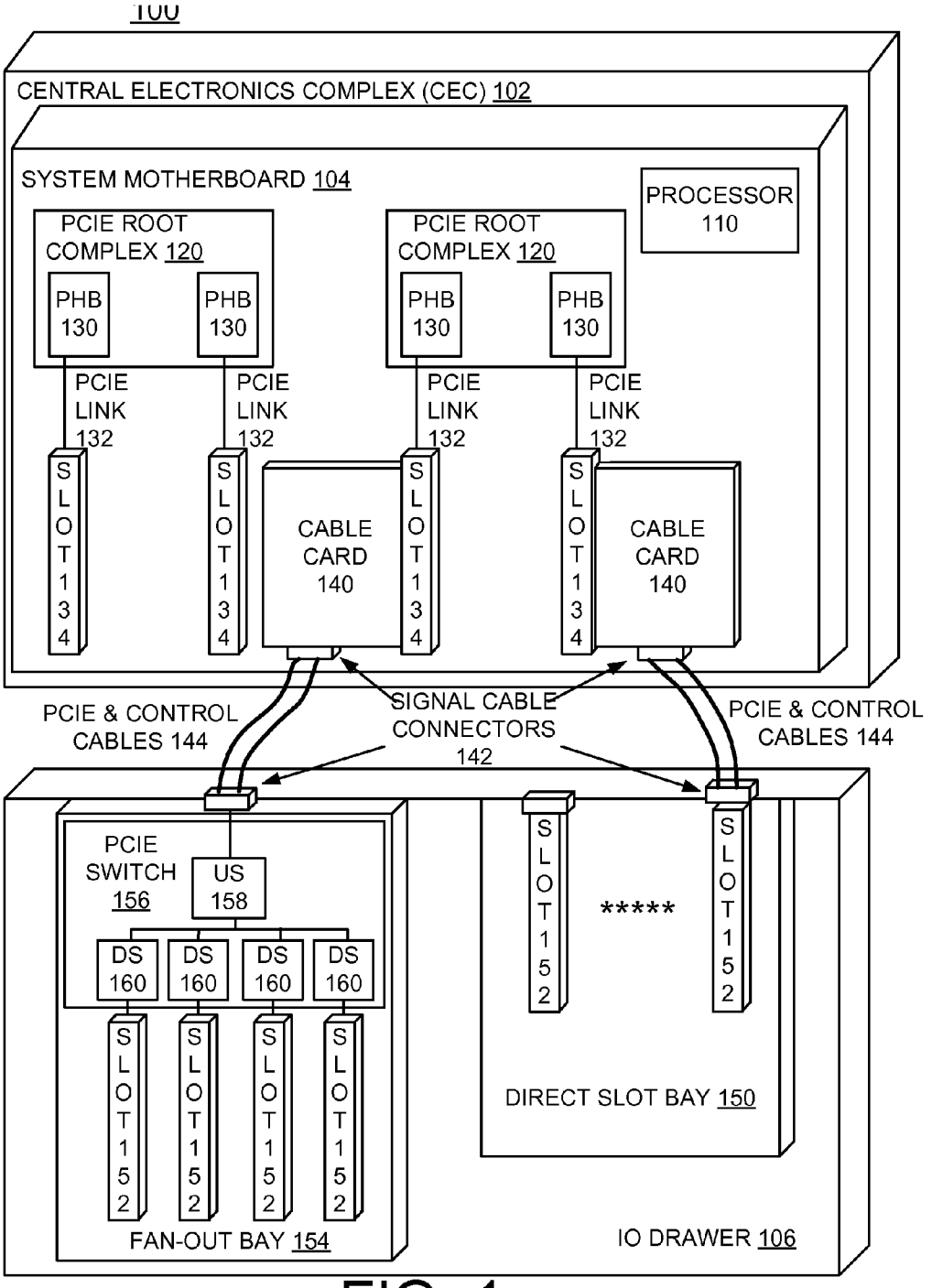

100
CENTRAL ELECTRONICS COMPLEX (CEC) 102
SYSTEM MOTHERBOARD 104
PROCESSOR 110
PCIE ROOT COMPLEX 120
PCIE ROOT COMPLEX 120
PHB 130
PHB 130
PHB 130
PHB 130
PCIE LINK 132
PCIE LINK 132
PCIE LINK 132
PCIE LINK 132
SLOT 134
SLOT 134
SLOT 134
SLOT 134
CABLE CARD 140
CABLE CARD 140
PCIE & CONTROL CABLES 144
SIGNAL CABLE CONNECTORS 142
PCIE & CONTROL CABLES 144
PCIE SWITCH 156
US 158
DS 160
DS 160
DS 160
DS 160
SLOT 152
SLOT 152
SLOT 152
SLOT 152
FAN-OUT BAY 154
SLOT 152
*****
SLOT 152
DIRECT SLOT BAY 150
IO DRAWER 106

FIG. 1

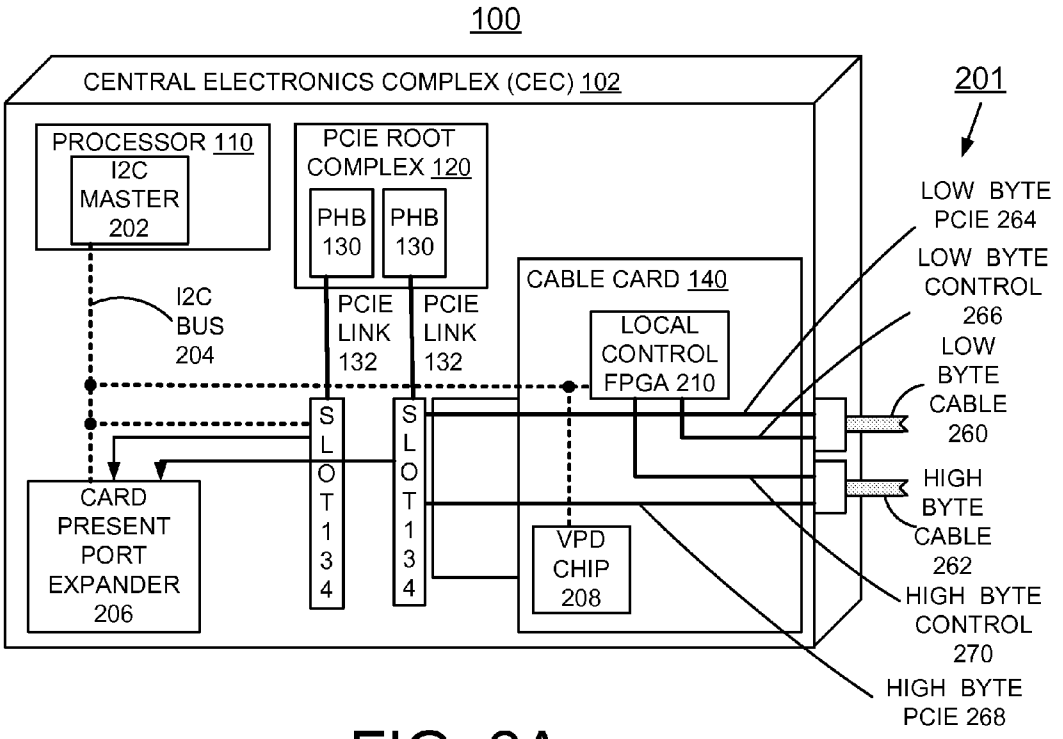

100
CENTRAL ELECTRONICS COMPLEX (CEC) 102
201
PROCESSOR 110
I2C MASTER 202
PCIE ROOT COMPLEX 120
PHB 130
PHB 130
CABLE CARD 140
LOCAL CONTROL FPGA 210
I2C BUS 204
PCIE LINK 132
PCIE LINK 132
SLOT 134
SLOT 134
CARD PRESENT PORT EXPANDER 206
VPD CHIP 208
LOW BYTE PCIE 264
LOW BYTE CONTROL 266
LOW BYTE CABLE 260
HIGH BYTE CABLE 262
HIGH BYTE CONTROL 270
HIGH BYTE PCIE 268

FIG. 2A

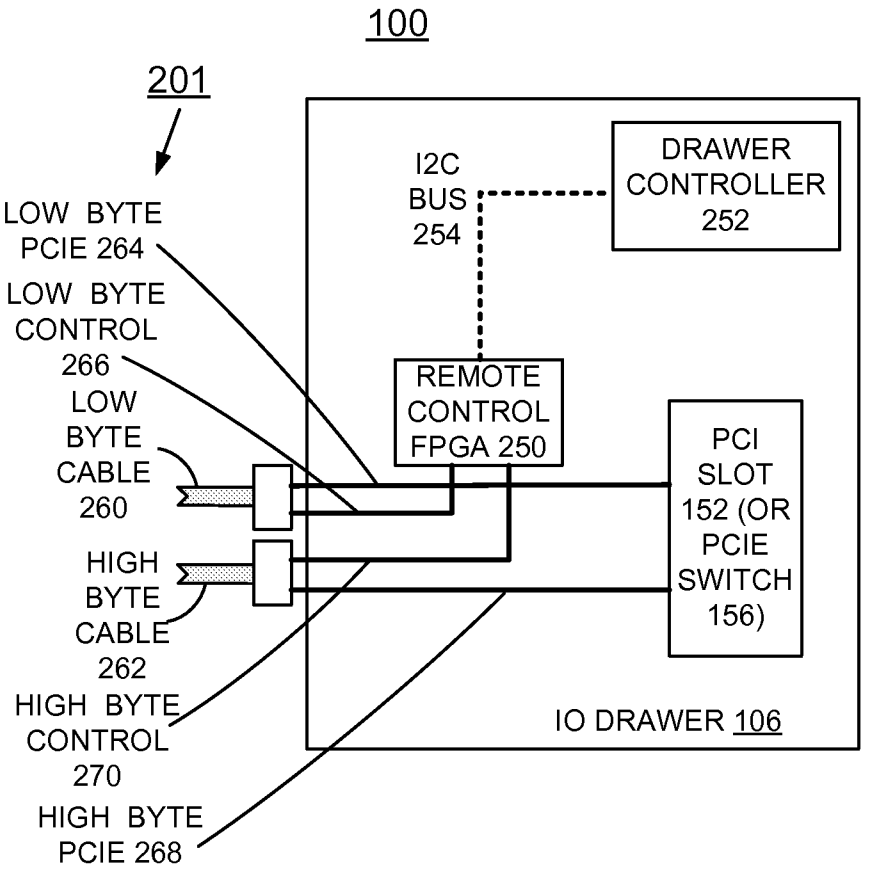

100
201
LOW BYTE PCIE 264
LOW BYTE CONTROL 266
LOW BYTE CABLE 260
HIGH BYTE CABLE 262
HIGH BYTE CONTROL 270
HIGH BYTE PCIE 268
I2C BUS 254
DRAWER CONTROLLER 252
REMOTE CONTROL FPGA 250
PCI SLOT 152 (OR PCIE SWITCH 156)
IO DRAWER 106

FIG. 2B

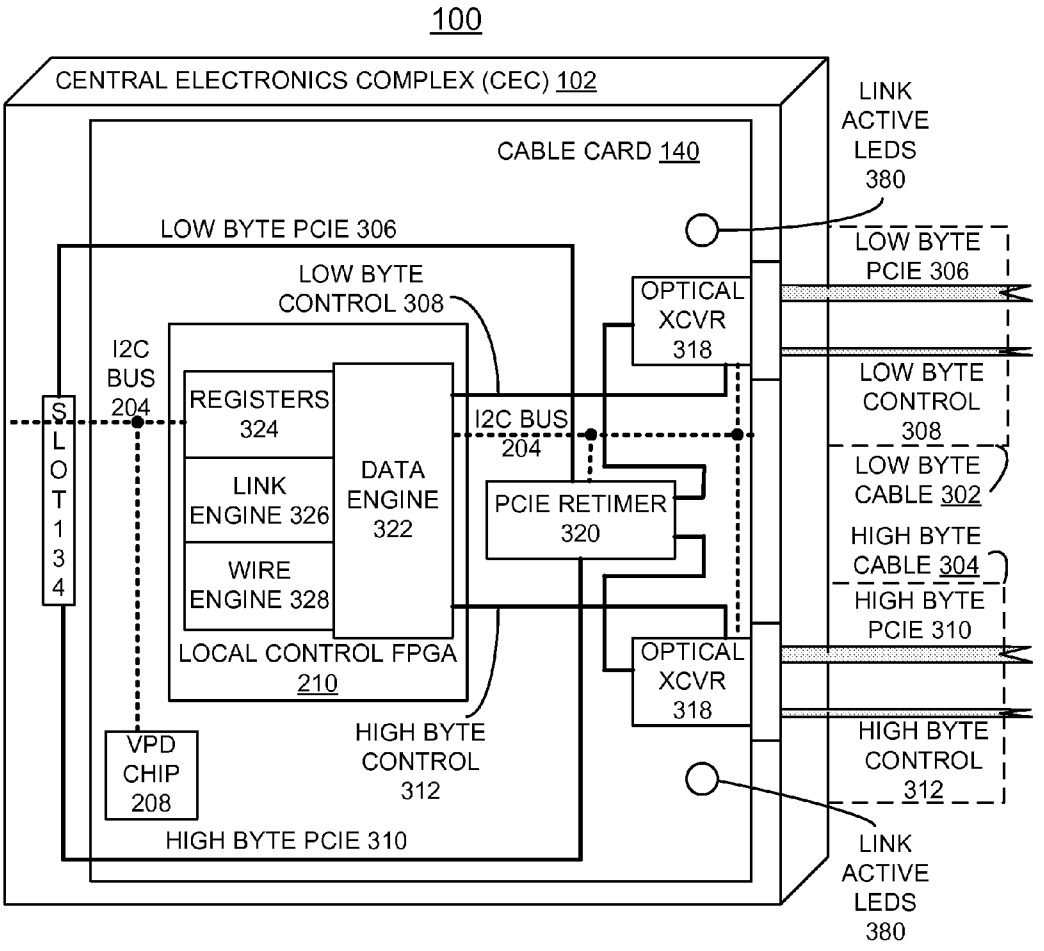

100
CENTRAL ELECTRONICS COMPLEX (CEC) 102
CABLE CARD 140
LOW BYTE PCIE 306
LOW BYTE CONTROL 308
I2C BUS 204
SLOT 134
REGISTERS 324
LINK ENGINE 326
WIRE ENGINE 328
DATA ENGINE 322
LOCAL CONTROL FPGA 210
VPD CHIP 208
HIGH BYTE CONTROL 312
HIGH BYTE PCIE 310
I2C BUS 204
PCIE RETIMER 320
OPTICAL XCVR 318
OPTICAL XCVR 318
LINK ACTIVE LEDS 380
LOW BYTE PCIE 306
LOW BYTE CONTROL 308
LOW BYTE CABLE 302
HIGH BYTE CABLE 304
HIGH BYTE PCIE 310
HIGH BYTE CONTROL 312
LINK ACTIVE LEDS 380

FIG. 3A

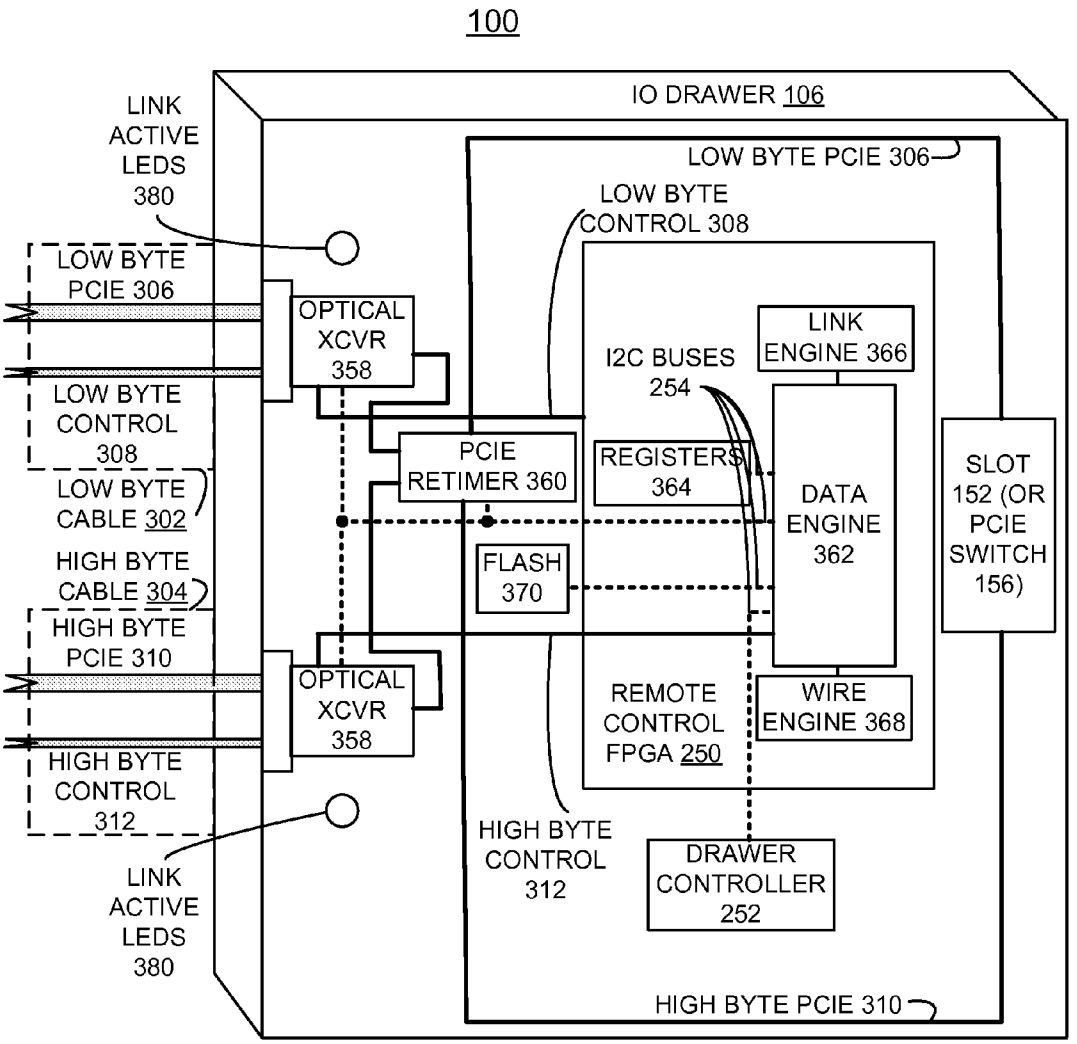

100
IO DRAWER 106
LINK ACTIVE LEDS 380
LOW BYTE PCIE 306
LOW BYTE CONTROL 308
LOW BYTE CABLE 302
HIGH BYTE CABLE 304
HIGH BYTE PCIE 310
HIGH BYTE CONTROL 312
LINK ACTIVE LEDS 380
OPTICAL XCVR 358
OPTICAL XCVR 358
PCIE RETIMER 360
FLASH 370
I2C BUSES 254
REGISTERS 364
LINK ENGINE 366
DATA ENGINE 362
WIRE ENGINE 368
REMOTE CONTROL FPGA 250
DRAWER CONTROLLER 252
SLOT 152 (OR PCIE SWITCH 156)
LOW BYTE PCIE 306
LOW BYTE CONTROL 308
HIGH BYTE CONTROL 312
HIGH BYTE PCIE 310

FIG. 3B

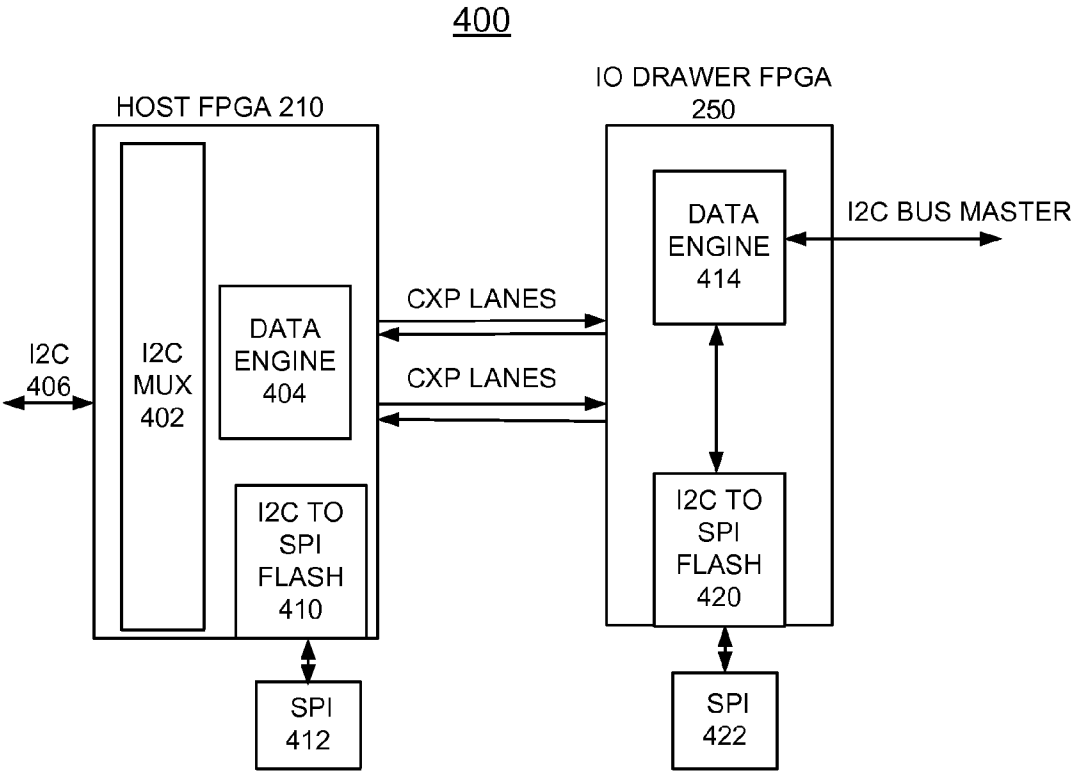

400
HOST FPGA 210
IO DRAWER FPGA 250
I2C 406
I2C MUX 402
DATA ENGINE 404
CXP LANES
CXP LANES
DATA ENGINE 414
I2C BUS MASTER
I2C TO SPI FLASH 410
I2C TO SPI FLASH 420
SPI 412
SPI 422

FIG. 4

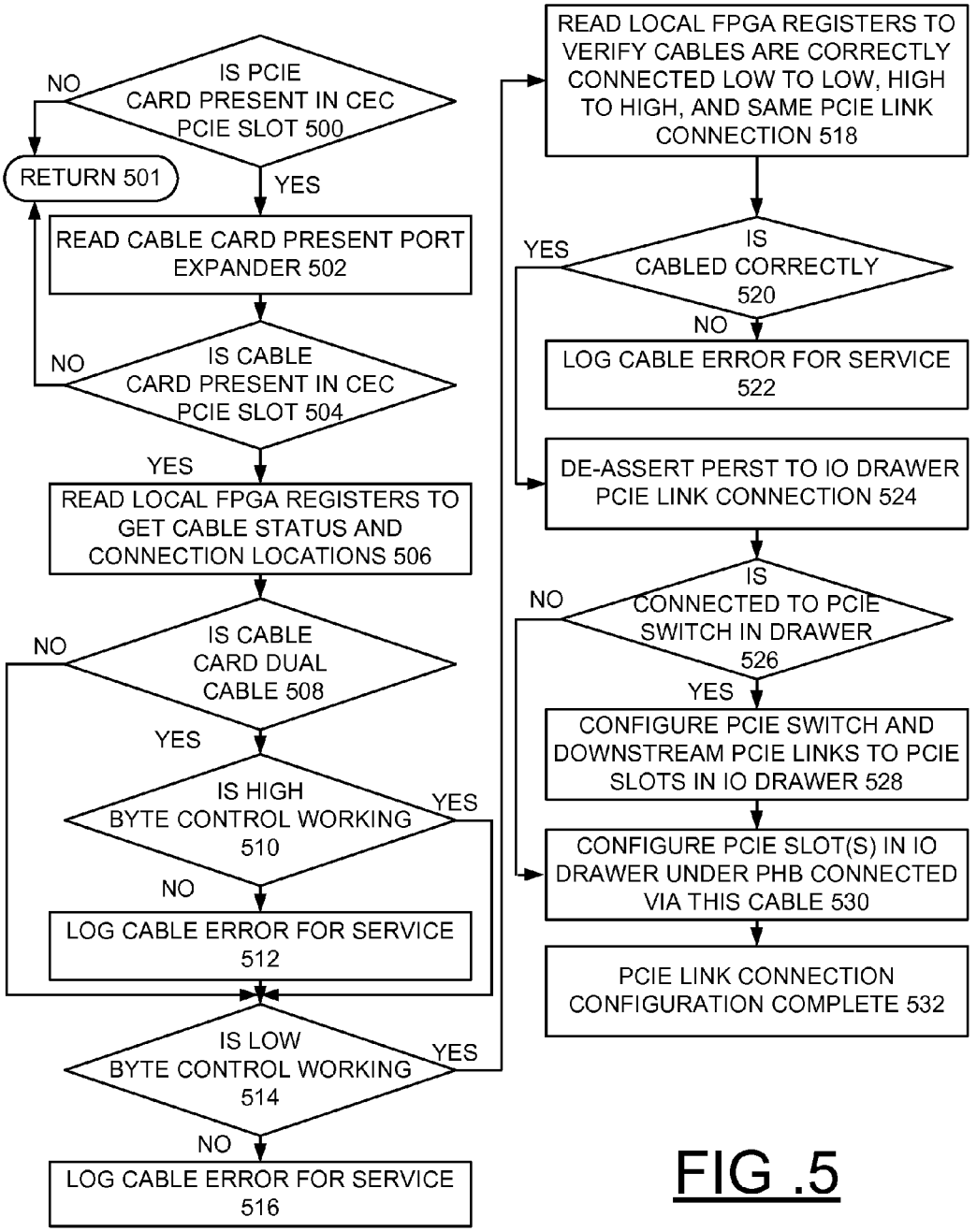

IS PCIE CARD PRESENT IN CEC PCIE SLOT 500
NO
YES
RETURN 501
READ CABLE CARD PRESENT PORT EXPANDER 502
IS CABLE CARD PRESENT IN CEC PCIE SLOT 504
NO
YES
READ LOCAL FPGA REGISTERS TO GET CABLE STATUS AND CONNECTION LOCATIONS 506
IS CABLE CARD DUAL CABLE 508
NO
YES
IS HIGH BYTE CONTROL WORKING 510
YES
NO
LOG CABLE ERROR FOR SERVICE 512
IS LOW BYTE CONTROL WORKING 514
YES
NO
LOG CABLE ERROR FOR SERVICE 516
READ LOCAL FPGA REGISTERS TO VERIFY CABLES ARE CORRECTLY CONNECTED LOW TO LOW, HIGH TO HIGH, AND SAME PCIE LINK CONNECTION 518
IS CABLED CORRECTLY 520
YES
NO
LOG CABLE ERROR FOR SERVICE 522
DE-ASSERT PERST TO IO DRAWER PCIE LINK CONNECTION 524
IS CONNECTED TO PCIE SWITCH IN DRAWER 526
NO
YES
CONFIGURE PCIE SWITCH AND DOWNSTREAM PCIE LINKS TO PCIE SLOTS IN IO DRAWER 528
CONFIGURE PCIE SLOT(S) IN IO DRAWER UNDER PHB CONNECTED VIA THIS CABLE 530
PCIE LINK CONNECTION CONFIGURATION COMPLETE 532

FIG .5

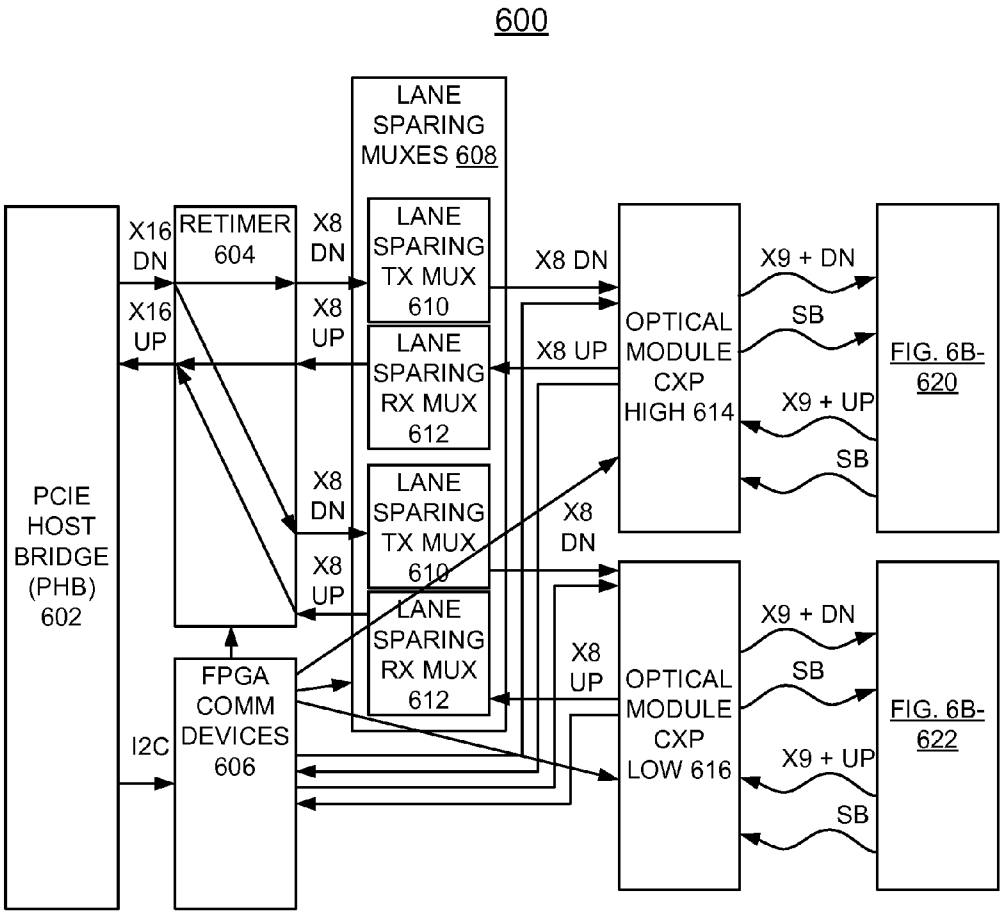

600
LANE SPARING MUXES 608
PCIE HOST BRIDGE (PHB) 602
X16 DN
X16 UP
RETIMER 604
X8 DN
X8 UP
LANE SPARING TX MUX 610
LANE SPARING RX MUX 612
X8 DN
X8 UP
LANE SPARING TX MUX 610
LANE SPARING RX MUX 612
FPGA COMM DEVICES 606
I2C
X8 DN
X8 UP
X8 DN
X8 UP
OPTICAL MODULE CXP HIGH 614
OPTICAL MODULE CXP LOW 616
X9 + DN
SB
X9 + UP
SB
X9 + DN
SB
X9 + UP
SB
FIG. 6B-620
FIG. 6B-622

FIG. 6A

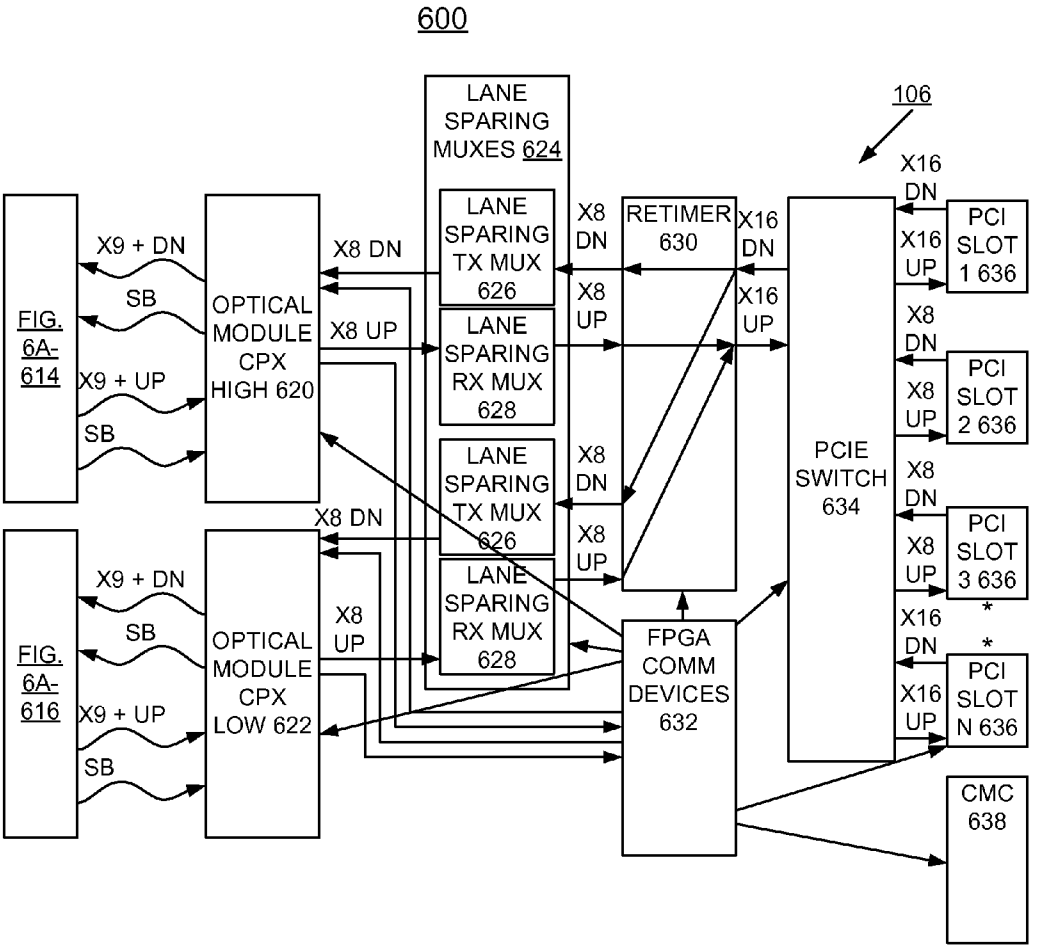

600
106
LANE SPARING MUXES 624
LANE SPARING TX MUX 626
LANE SPARING RX MUX 628
LANE SPARING TX MUX 626
LANE SPARING RX MUX 628
RETIMER 630
FPGA COMM DEVICES 632
PCIE SWITCH 634
FIG. 6A-614
FIG. 6A-616
OPTICAL MODULE CPX HIGH 620
OPTICAL MODULE CPX LOW 622
X9 + DN
SB
X9 + UP
SB
X9 + DN
SB
X9 + UP
SB
X8 DN
X8 UP
X8 DN
X8 UP
X8 DN
X8 UP
X8 DN
X8 UP
X16 DN
X16 UP
X16 DN
X16 UP
X16 DN
X16 UP
X8 DN
X8 UP
X8 DN
X8 UP
X8 DN
X8 UP
X16 DN
X16 UP
PCI SLOT 1 636
PCI SLOT 2 636
PCI SLOT 3 636
PCI SLOT N 636
CMC 638

FIG. 6B

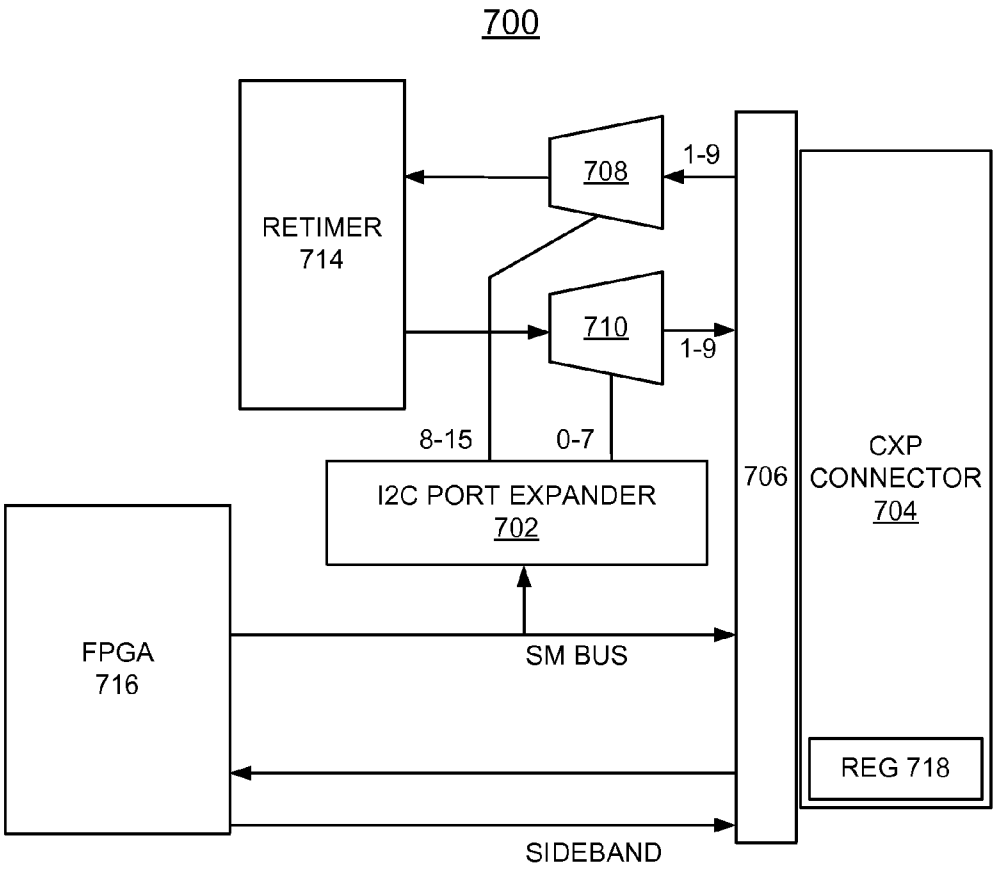

700
RETIMER
714
708
1-9
710
1-9
8-15
0-7
I2C PORT EXPANDER
702
706
CXP
CONNECTOR
704
FPGA
716
SM BUS
REG 718
SIDEBAND

FIG. 7

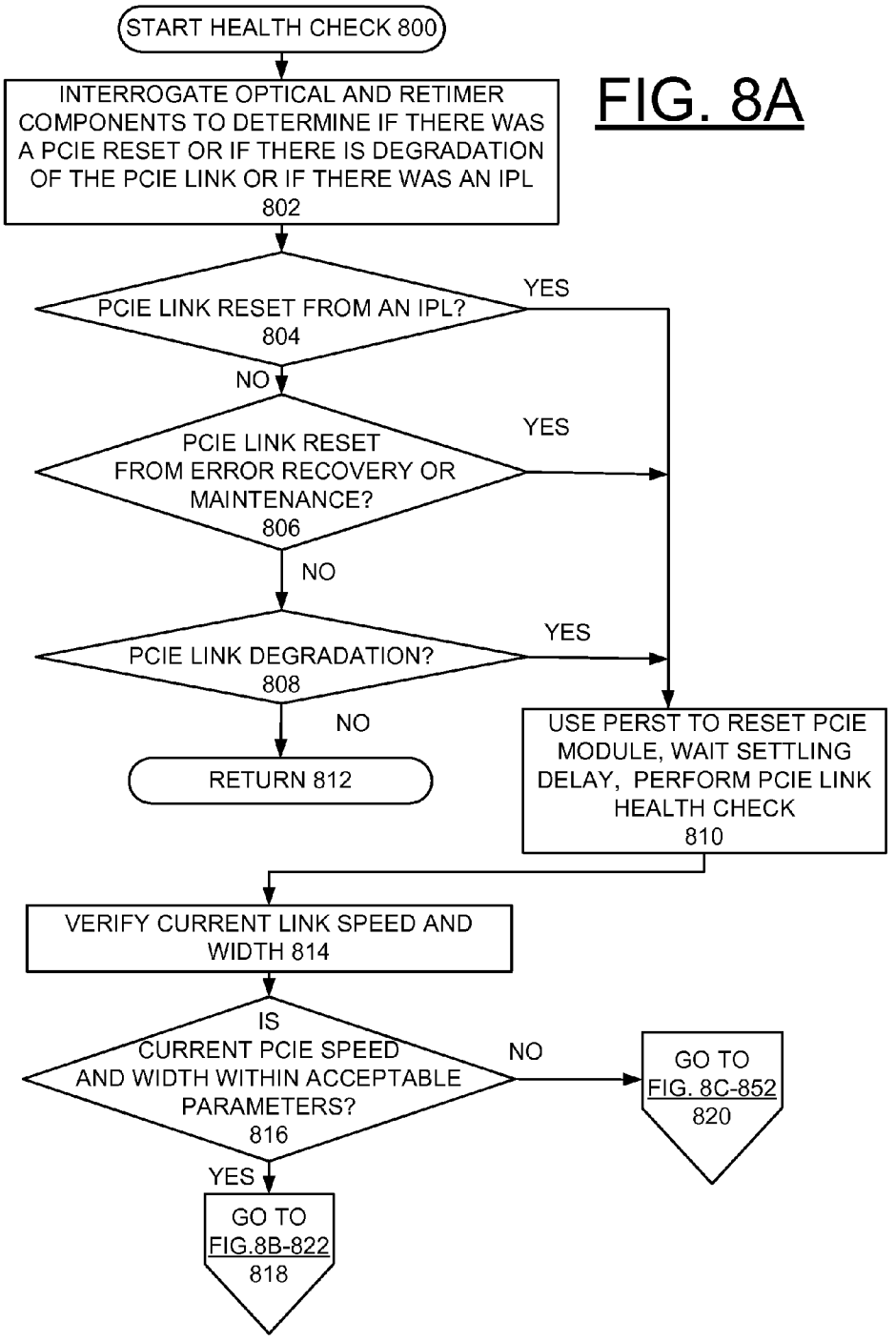

START HEALTH CHECK 800
INTERROGATE OPTICAL AND RETIMER COMPONENTS TO DETERMINE IF THERE WAS A PCIE RESET OR IF THERE IS DEGRADATION OF THE PCIE LINK OR IF THERE WAS AN IPL
802
PCIE LINK RESET FROM AN IPL?
804
YES
NO
PCIE LINK RESET FROM ERROR RECOVERY OR MAINTENANCE?
806
YES
NO
PCIE LINK DEGRADATION?
808
YES
NO
RETURN 812
USE PERST TO RESET PCIE MODULE, WAIT SETTLING DELAY, PERFORM PCIE LINK HEALTH CHECK
810
VERIFY CURRENT LINK SPEED AND WIDTH 814
IS CURRENT PCIE SPEED AND WIDTH WITHIN ACCEPTABLE PARAMETERS?
816
NO
GO TO FIG. 8C-852
820
YES
GO TO FIG.8B-822
818

FIG. 8A

FIG. 8B

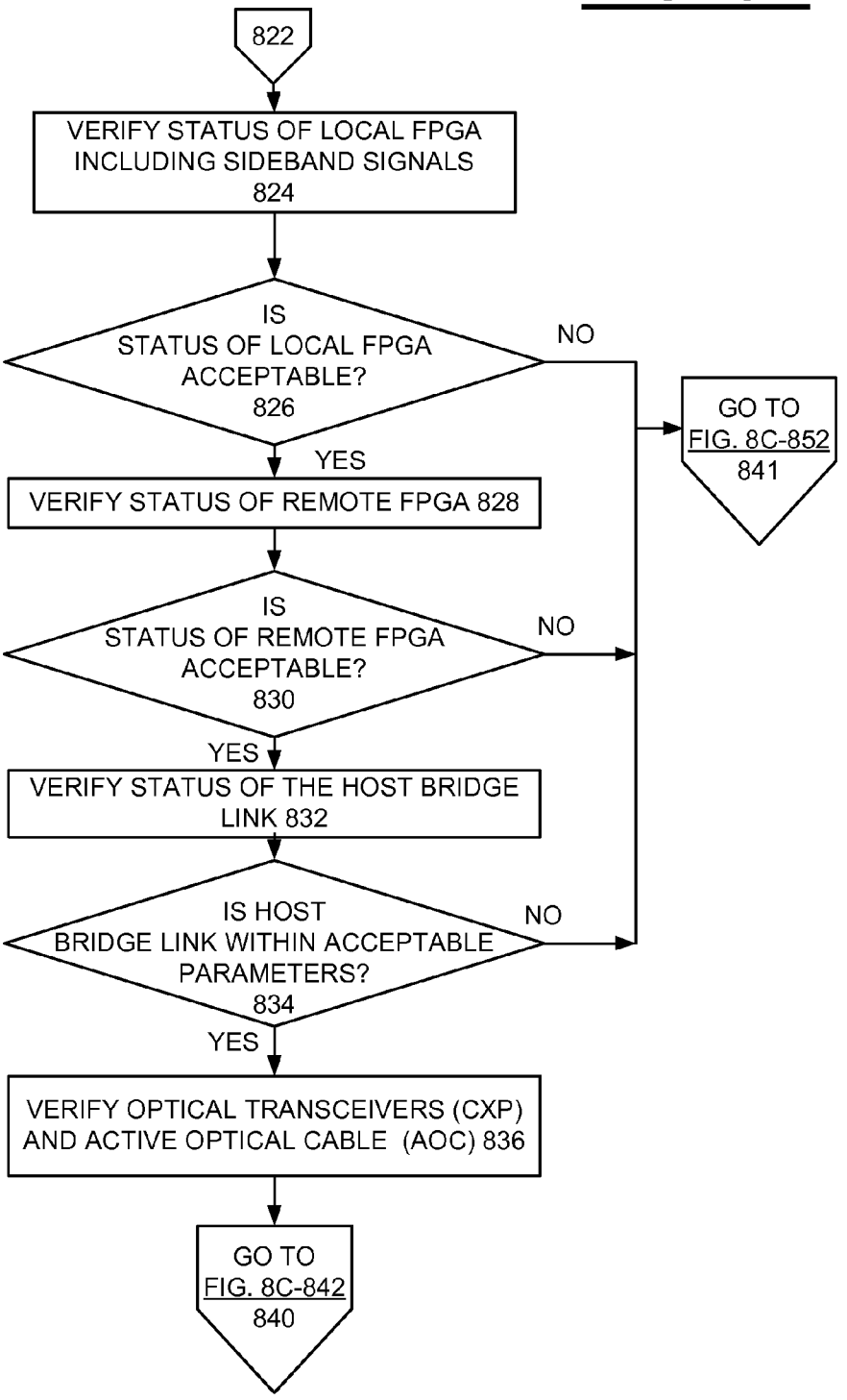
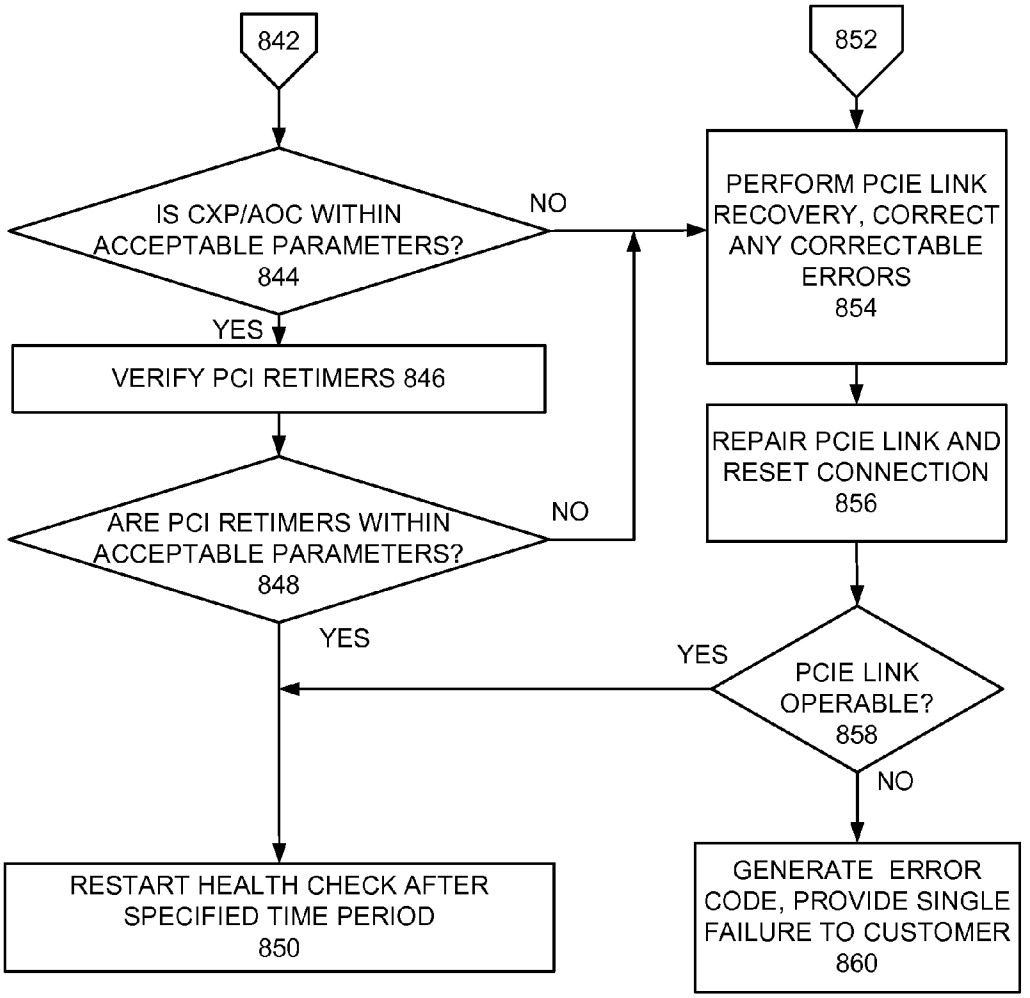

822
VERIFY STATUS OF LOCAL FPGA INCLUDING SIDEBAND SIGNALS 824
IS STATUS OF LOCAL FPGA ACCEPTABLE? 826
NO
GO TO FIG. 8C-852 841
YES
VERIFY STATUS OF REMOTE FPGA 828
IS STATUS OF REMOTE FPGA ACCEPTABLE? 830
NO
YES
VERIFY STATUS OF THE HOST BRIDGE LINK 832
IS HOST BRIDGE LINK WITHIN ACCEPTABLE PARAMETERS? 834
NO
YES
VERIFY OPTICAL TRANSCEIVERS (CXP) AND ACTIVE OPTICAL CABLE (AOC) 836
GO TO FIG. 8C-842 840

FIG. 8C


842
IS CXP/AOC WITHIN ACCEPTABLE PARAMETERS? 844
NO
YES
VERIFY PCI RETIMERS 846
ARE PCI RETIMERS WITHIN ACCEPTABLE PARAMETERS? 848
NO
YES
RESTART HEALTH CHECK AFTER SPECIFIED TIME PERIOD 850
852
PERFORM PCIE LINK RECOVERY, CORRECT ANY CORRECTABLE ERRORS 854
REPAIR PCIE LINK AND RESET CONNECTION 856
PCIE LINK OPERABLE? 858
YES
NO
GENERATE ERROR CODE, PROVIDE SINGLE FAILURE TO CUSTOMER 860

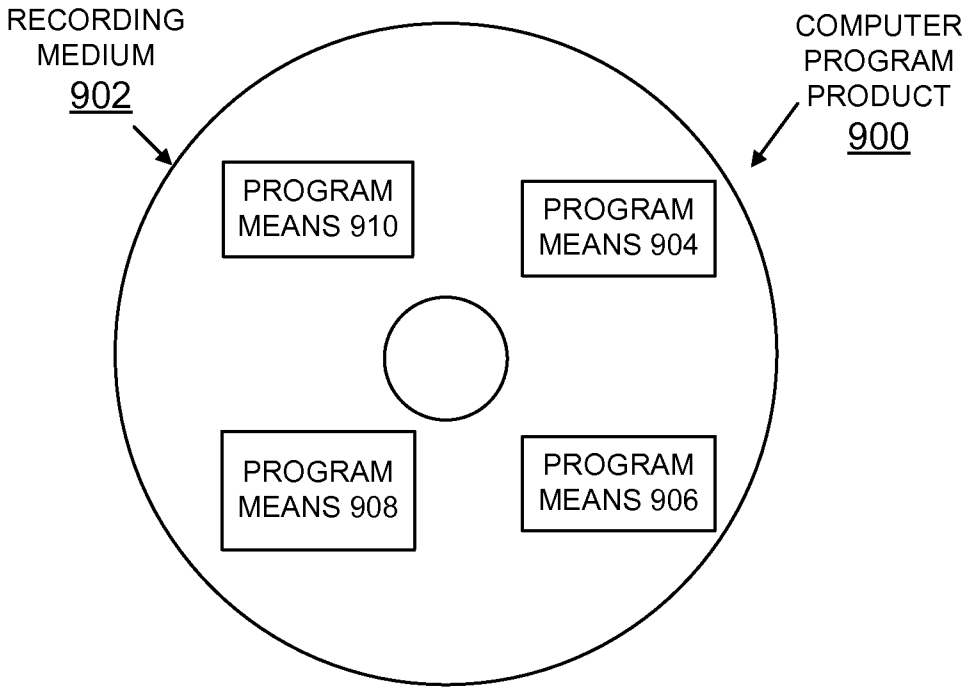

RECORDING MEDIUM
902
COMPUTER PROGRAM PRODUCT
900
PROGRAM MEANS 910
PROGRAM MEANS 904
PROGRAM MEANS 908
PROGRAM MEANS 906

FIG. 9

# IMPLEMENTING HEALTH CHECK FOR OPTICAL CABLE ATTACHED PCIE ENCLOSURE

## FIELD OF THE INVENTION

**[0001]** The present invention relates generally to the data processing field, and more particularly, relates to a method, system and computer program product for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCI-Express or PCIE) enclosures in a computer system.

## DESCRIPTION OF THE RELATED ART

**[0002]** Peripheral Component Interconnect Express (PCIE) has become the industry standard IO bus for server computer systems, as well as personal computers (PCs). Traditionally, servers install PCIE IO adapters (IOAs) in slots within a system unit that connect through a PCI host bridge to the system memory and processor buses. IBM POWER and Z series systems have offered external IO enclosures to provide additional PCIE slots beyond those that are available within the system unit. These have in the past been connected to the system unit through IBM proprietary interconnect architectures such as HSL and Infiniband 12X IO loops on IBM POWER systems.

**[0003]** In PCIE fabrics, it is important for system performance to operate the connections between devices and the host (PCIE links) at optimal bandwidth (speed and width). A multi component solution for attaching an external PCIE enclosure to a PCIE host presents additional challenges, with longer signal lengths and/or active components in the path creating a greater possibility for signal integrity issues. PCIE allows devices forming a link to automatically reduce the speed or width of the connection when the link cannot operate reliably at a higher speed. For example, the link may be trained and functional at 8 GT/s, then issues are detected and the link goes into recovery and speed changes to 5 GT/s. Many PCIE Root Complexes do not provide a mechanism for notifying a host of a link that has dropped in width or speed from the reported maximum. Notification is only provided if the PCI Express link drops through the assertion of a link down interrupt.

**[0004]** A need exists to provide an effective mechanism for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCI-Express or PCIE) enclosures in a computer system.

## SUMMARY OF THE INVENTION

**[0005]** Principal aspects of the present invention are to provide a method, system and computer program product for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCI-Express or PCIE) enclosures in a computer system. Other important aspects of the present invention are to provide such method, system and computer program product substantially without negative effects and that overcome many of the disadvantages of prior art arrangements.

**[0006]** In brief, a method, system and computer program product are provided for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCIE) enclosures in a computer system. System firmware is provided for implementing health check functions. One or more optical cables are connected between a host bridge and a PCIE enclosure. A PCIE link to the PCIE enclosure is reset responsive to a predefined event. After a settling delay, a PCIE link health check is performed verifying PCIE link width and speed.

**[0007]** In accordance with features of the invention, the PCIE link is reset and retrained from an initial program load (IPL) of the system, a recovery from an error and as a result of a concurrent maintenance operation. A predefined control signal is used to reset the attached PCIE enclosure.

**[0008]** In accordance with features of the invention, the PCIE link health check is performed at key points triggered by both correctable and uncorrectable failure where a likely cause is a link degradation.

**[0009]** In accordance with features of the invention, the polled health check and the predicative health check work together to prevent multiple detections of link degradation and to facilitate a clean recovery.

**[0010]** In accordance with features of the invention, when a degraded PCIE link is found, a recovery procedure is performed to attempt to bring the link back to its full optimal width and speed.

**[0011]** In accordance with features of the invention, the polled health check and the predicative health check of the optical link includes optical transceivers, active fibre or active optical cable (AOC) verified to be running at its full PCIE link width and speed.

**[0012]** In accordance with features of the invention, the health check includes verifying the current link speed and width and comparing to the reported speed and width. The health check includes verifying status of a local field programmable gate array (FPGA) including the sideband signals. The health check includes verifying status of a remote field programmable gate array (FPGA). The health check includes verifying status of a Host Bridge Link (HBL). The health check includes verifying status of optical transceivers, and active optical cable (AOC). The health check includes verifying status of PCI retimers.

**[0013]** In accordance with features of the invention, the health check is restarted after a specified time period. For example, a process is started that initiates the health check at one minute intervals.

**[0014]** In accordance with features of the invention, the PCIE health check is run during the recovery of errors related to the PCI link.

**[0015]** In accordance with features of the invention, the health check is executed during the recovery of PCIE correctable errors, PCIE uncorrectable errors and PHB link related errors including link down. If the health check discovers a correctable active optical cable (AOC) error the link will be repaired and the connection reset. The AOC error takes precedence over an original reported error.

**[0016]** In accordance with features of the invention, if the health check detects a PCIE link degradation, the health check delays and rechecks the link after a specified time period such as 1 minute, allowing any errors in the platform to be discovered first and initiate the recovery.

**[0017]** In accordance with features of the invention, the PCIE health check is run during an initial IPL of the system ensuring the links are running at their maximum capability.

**[0018]** In accordance with features of the invention, an Error Identification (ID) related to the health check is correlated with Error ID generated from error handling code for the Root Complex PCIE Host Bridge (PHB) or the first upstream device in the enclosure to provide a single failure to the

customer. When the health check is initiated by an error recovery mechanism for the PHB or the PCIE Switch, and the health check finds an issue in the link path, it will be ensured that errors that are reported are all properly correlated with each other, resulting in one error ID in the logs. Error IDs will be propagated from the error recovery handlers to the health check, to ensure these errors stay correlated.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0019]** The present invention together with the above and other objects and advantages may best be understood from the following detailed description of the preferred embodiments of the invention illustrated in the drawings, wherein:

**[0020]** FIG. **1** illustrates an example computer system for implementing sideband control structure for Peripheral Component Interconnect Express (PCIe or PCI-Express) PCIe cable cards and input/output (IO) expansion enclosures and health check for optical cable attached PCIE enclosures in accordance with a preferred embodiment;

**[0021]** FIGS. **2**A, and **2**B illustrates cable card and sideband controls of the example system of FIG. **1** in accordance with a preferred embodiment;

**[0022]** FIGS. **3**A, and **3**B illustrates further details of Local and Remote FPGA Controllers of the example system of FIG. **1** in accordance with a preferred embodiment;

**[0023]** FIG. **4** illustrates example operational features of a physical layer view of I2C multiplexer (MUX) and data engine at host side and IO drawer end of the example system of FIG. **1** in accordance with a preferred embodiment;

**[0024]** FIG. **5** is a flow chart illustrating example firmware operational features using sideband status information of the example system of FIG. **1** in accordance with a preferred embodiment;

**[0025]** FIGS. **6**A and **6**B together illustrate example PCI bus topology of the example system of FIG. **1** in accordance with a preferred embodiment;

**[0026]** FIG. **7** illustrate example lane sparing hardware of the example system of FIG. **1** in accordance with a preferred embodiment;

**[0027]** FIGS. **8**A, **8**B, and **8**C together provide a flow chart illustrating example operations for implementing heath check for optical cable attached PCIE enclosures in accordance with a preferred embodiment; and

**[0028]** FIG. **9** is a block diagram illustrating a computer program product in accordance with the preferred embodiment.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

**[0029]** In the following detailed description of embodiments of the invention, reference is made to the accompanying drawings, which illustrate example embodiments by which the invention may be practiced. It is to be understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the invention.

**[0030]** The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

**[0031]** In accordance with features of the invention, a method, system and computer program product are provided for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCI-Express or PCIE) enclosures in a computer system. Following a reset between a host bridge to the PCIE enclosure over the active optical cable, a PCIE link is trained. The PCIE link is reset and retrained, for example, from an initial program load (IPL) of the system, a recovery from an error and as a result of a concurrent maintenance operation. A predefined control signal is used to reset the attached PCIE enclosure, after a settling delay, the PCIE link health check is performed verifying PCIE link width and speed. The PCIE link health check is performed at key points triggered by both correctable and uncorrectable failure where a likely cause is a link degradation. Polled health check and predicative health check work with each other to prevent multiple detections of link degradation and to facilitate a clean recovery. When a degraded PCIE link is found, a recovery procedure is performed to attempt to bring the link back to its full optimal width and speed.

**[0032]** Having reference now to the drawings, in FIG. **1**, there is shown an example computer system generally designated by the reference character **100** for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCI-Express or PCIE) enclosures and implementing sideband control structure for Peripheral Component Interconnect Express (PCIE or PCI-Express) PCIE cable cards and input/output (IO) expansion enclosures in accordance with a preferred embodiment. Computer system **100** includes a computer Central Electronics Complex (CEC) **102** that includes a system motherboard **104** common in the art that provides for the physical packaging interconnection of the various components illustrated within the CEC **102**. Computer system **100** includes an input/output (IO) expansion enclosure or IO drawer generally designated by the reference character **106** in accordance with a preferred embodiment.

**[0033]** In accordance with features of the invention, the IO drawer **106** is an electrical enclosure that provides PCIE add-in card slots (PCIE slots) or integrated devices utilized by firmware and software operating within the CEC **102** to extend the number of PCIE devices physically configurable within the scope of the CEC.

**[0034]** CEC system motherboard **104** includes one or more processors or central processor units (CPUs), such as processor **110**. Processor **110** is suitably arranged for executing firmware and software, such as operating system and applications hosted by respective operating systems in control of the overall computer system **100** in accordance with a preferred embodiment.

**[0035]** CEC system motherboard **104** includes one or a plurality of PCIE root complexes **120**. Each PCIE root complex **120** includes one or a plurality of PCIE host bridges (PHBs) **130**, each PHB **130** providing a respective PCIE link **132** to a respective PCIE slot **134** mounted commonly on the system motherboard **104**. For example, the PCIE link **132** is comprised of a plurality of 16 lanes of PCIE signaling wires and auxiliary signals, such as specified in the PCIE standard. The auxiliary signals include PCIE PERST, also called Fun-

damental Reset, and PCIE clocks from the PHB **130** to a PCIE device, PCIE card present from a device to a PHB **130**, and other such discrete signals that operate independently of the PCIE signaling lanes.

**[0036]** One or more of the PCIE slots **134** includes a respective PCIE cable card **140** including a respective PCIE cable connector **142** connected by a pair of PCIE cables **144** to a corresponding respective PCIE cable connector **142** provided with the IO drawer **106**, as shown.

**[0037]** IO drawer **106** includes one or a plurality of module bays, such as a direct slot bay **150** including a plurality of PCIE slots **152**, and a fan-out bay **154**, including a PCIE switch **156** connected to another plurality of PCIE slots **152**, as shown. Each of the direct slot bay **150**, PCIE slots **152**, and the fan-out bay **154** is connected to the CEC **102** by the individual and independent cables **144** via PCIE cable connectors **142**, as shown. The PCIE switch **156** includes an upstream port **158** connected upstream to the PCIE Cable Connector **142** and downstream to a plurality of downstream ports **160**, each connected to a respective PCIE slot **152**, as shown.

**[0038]** Computer system **100** is shown in simplified form sufficient for understanding the present invention. The illustrated computer system **100** is not intended to imply architectural or functional limitations. The present invention can be used with various hardware implementations and systems and various other internal hardware devices.

**[0039]** In accordance with features of the invention, one or a plurality of cables **144** is provided with each cable card **140** and respective PCIE slot **152** or PCIE switch **156**. As shown, a pair of PCIE cables **144** convey PCIE signal lanes and auxiliary signals forming a single PCIE link, and the PCIE cables **144** also convey sideband control and status signals between the CEC **102** and the IO drawer **106**. The use of two cables **144** for each single PCIE link between the CEC **102** and IO drawer **106** advantageously enables redundancy in the event of a cable failure. Only a single cable **144** is utilized in communicating sideband controls and status between the CEC and IO drawer, and each of the two cables **144** provides an alternative signaling path to communicate such sideband controls.

**[0040]** In accordance with features of the invention, the cables **144** may be conventional copper cables or fiber optic cables employing opto-electronic transceivers at each cable end. For example, with relatively short distances, such as inches to a few feet, copper cables **144** are generally suitable for conveying PCIE lanes operating at required PCIE signaling rates between the CEC **102** and IO drawer **106**. Optical cables **144** advantageously provide electrical isolation to enable high signaling rates over longer distances, such as one to many meters.

**[0041]** In accordance with features of the invention, use of optical PCIE cables **144** includes serializing required DC signals over the optical cables **144** and then converting the serialized signals back to DC signals at the other end of the optical PCIE cables **144** at CEC **102** or the IO drawer **106**.

**[0042]** In accordance with features of the invention, firmware operating in the CEC **102** is enabled to determine a type of cable card **140** plugged into a PCIE slot **134**, with a plurality of different such cable card types having differing properties, such as whether the cables **144** are copper or optical, how many cables **144** emanate from the cable card **140**, and the like. It is similarly advantageous for firmware operating in the CEC **102** to be able to determine whether the cables **144** are actually connected to an IO drawer **106**, and what type of IO drawer **106**, and the IO module bay **150**, or module bay **156**, is connected to cables **144**.

**[0043]** In accordance with features of the invention, in a logically partitioned computer system **100** firmware operating in the CEC **102** is enabled to assign respective CEC PCIE slots **134** to a particular logical partition when that slot **134** contains a PCIE IO adapter or firmware operating in the CEC **102** is enabled to detect a cable card **140** and to assign respective CEC PCIE slots **152** to different logical partitions with the cable card **140** connected to an IO drawer **106**.

**[0044]** Referring also to FIGS. **2**A, and **2**B, additional example details of novel cable card **140** and sideband controls generally designated by the reference character **201** are shown of computer system **100** of FIG. **1** in accordance with a preferred embodiment. As shown in FIG. **2**A, included within or accessible to the processor **110** is an Inter-Integrated Circuit (I2C) master device **202**, which is the master of an I2C bus **204** utilized as an IO bus of the sideband signaling apparatus of a preferred embodiment. It should be understood that other such IO buses known in the art may be suitable to substitute for the I2C bus **204** utilized by the invention. Within CEC **102**, the I2C bus **204** is connected between the I2C master device **202** and a card present port expander **206**, PCIE slots **134**, a vital product data (VPD) chip **208**, and a local control field programmable gate array (FPGA) **210** provided within the cable card **140**.

**[0045]** The PCIE cable card **140** utilizes pins within the PCIE signal connector **142** of the PCIE slot **134** defined in PCIE as reserved pins to generate a signal identifying the PCIE cable card **140** as a cable card. The card present port expander **206** connected on the I2C bus **204** receives a card present signal from the cable card **140** uniquely indicating the presence of a cable card, as opposed to a PCIE IO adapter. Firmware operating in the CEC **102** utilizes the I2C master **202** to read registers within the card present port expander **206** in order to determine that the cable card **140** is plugged in the respective PCIE card slot **134**. It should be understood that other devices than the card present port expander **206** could be used to receive cable card present information in a manner accessible to firmware operating within the CEC **102**.

**[0046]** The local control FPGA **210** includes registers that receive status from and optionally signal controls to other hardware components located on the cable card **140**. The registers within the local control FPGA **210** are connected to the I2C bus **204** proceeding from the PCIE slot **140** onto the cable card **140**.

**[0047]** Referring also to FIG. **2**B, the IO drawer **106** similarly includes a remote control FPGA **250**. The remote control FPGA **250** includes registers that receive signals from other hardware components internal to the IO drawer **106**. The IO drawer **106** includes a drawer controller **252** coupled to the remote control FPGA **250** via an I2C bus **254**.

**[0048]** In accordance with features of the invention, as shown in FIGS. **2**A, and **2**B, sideband controls **201** are coupled between the local control FPGA **210** in the CEC **102** and the remote control FPGA **250** in IO drawer **106** by a low byte cable **260** and a high byte cable **262**. For example, the low byte cable **260** conveys PCIE lanes 0 through 7 of the PCIE link from the PCIE slot **134**, shown as Low Byte PCIE **264**, and conveys sideband signals between the cable card **140** and the IO drawer **106**, shown as Low Byte Control **266**. For example, the high byte cable **262** conveys PCIE lanes 8 through 15 of the PCIE link from the PCIE slot **134**, shown as

High Byte PCIE **268**, and conveys sideband signals between the cable card **140** and the IO drawer **106**, shown as High Byte Control **270**. For example, the high byte cable **262** serves as an alternate or redundant connection to the low byte cable **260** for the purpose of conveying sideband signals.

**[0049]** Using either the low byte control **266** or high byte control **270**, the remote control FPGA **250** signals changes in the states of various hardware components or DC voltage signals within the IO drawer **106** to the local control FPGA **210**, which receives these changes in registers accessible to firmware operating in the CEC **102**. Similarly, firmware operating in the CEC **102** may set register values in the local control FPGA **210** directed at the remote control FPGA **250** to change the state of hardware components or DC voltage signals within the IO drawer **106**.

**[0050]** Using the either the low byte control **266** or high byte control **270**, the local FPGA **210** communicates local FPGA **210** register changes to the remote control FPGA **250**. The registers within the remote control FPGA **250** connect to the I2C bus **254** within the IO drawer **106**. The remote FPGA registers are also accessible as I2C devices from the local control FPGA **210**. Firmware operating in the CEC **102** utilizes registers in the local control FPGA **210** to create I2C bus operations transmitted between the local control FPGA **210** and remote control FPGA **250** utilizing the low byte control **266** or high byte control **270**. The local control FPGA **210** enables firmware operating within the CEC **102** to determine various configuration and operational states of hardware components or DC voltage signals located on the cable card **210** as well as hardware components or DC voltage signals within the IO drawer **106**.

**[0051]** The drawer controller **252** connected to the remote control FPGA **250** within the IO drawer **106** monitors or manages states of the hardware internal to the IO drawer, such as turning on or off power supplies within the drawer, monitoring thermal or electrical states of components within the drawer, taking actions in response to particular thermal or electrical states or thresholds, and the like. The drawer controller **252** connects to the remote control FPGA **250** utilizing the I2C bus **370**, enabling the drawer controller **252** to read or write registers within the remote control FPGA **250** and to communicate status to or receive control information communicated from the local control FPGA **210** using the low byte control **266** or high byte control **270**.

**[0052]** Referring also to FIGS. **3**A, and **3**B, additional example details of novel local control and remote control FPGAs **210**, **250** are shown of the example system **100** of FIG. **1** in accordance with a preferred embodiment.

**[0053]** The cable card **140** is shown connected to the IOA drawer **106** utilizing a low byte cable **302** and a high byte cable **304**. The low byte cable **302** conveys the low byte PCIE signals **306** representing PCIE lanes 0 to 7 and control signals between the local FPGA **210** and remote FPGA **250** indicated as low byte control **308**. The high byte cable **304** conveys the high byte PCIE signals **310** representing PCIE lanes 8 to 15 and control signals between the local FPGA **210** and remote FPGA **310** indicated as high byte control **312**. The signals conveyed by means of the low byte control **308** and high byte control **312** signals may be communicated over either or both of the low byte cable **302** and the high byte cable **304** at any one time, such that each cable can convey the control signals as a backup for the other in the event of failure or disconnection of one cable **302** or **304**, and such that signals may be communicated over both cables in order to detect the location to which each cable is connected at the IO drawer **106**.

**[0054]** The low byte control **308** and low byte PCIE **306** signals in the low byte cable **302**, and the high byte control **312** and high byte PCIE **310** signals in the high byte cable **304** are conveyed optically utilizing a respective optical transceiver (XCVR) **318** on the cable card **140** and optical transceiver (XCVR) **358** in the IO drawer **106**. The PCIE lanes 0 to 7 conveyed on the low byte PCIE **306** and lanes 8 to 15 conveyed on the high byte PCIE **310** commonly pass through a PCIE re-timer **320** in FIG. **3**A, a PCIE re-timer **360** in FIG. **3**A in order to synchronize them with the respective optical transceivers **318**, **358** shown in FIGS. **3**A and **3**B.

**[0055]** The local control FPGA **210** on the cable card **140** includes a local FPGA data engine **322**, registers **324**, a link engine **326**, and a wire engine **328**. The remote control FPGA **250** similarly includes a remote FPGA data engine **362**, registers **364**, a link engine **366**, and a wire engine **368**, and optionally a flash memory **270** coupled to the data engine **362**. The local control FPGA data engine **210** and remote control FPGA data engine **250** are capable of exchanging control signals utilizing either the low byte control **308** or high byte control **312** conveyed over the low byte cable **302** or high byte cable **304**, respectively. The link engine **326** utilizes the data engine **322** to establish reliable optical signaling and bit transfer protocols between the optical XCVRs **318** on the cable card **140** and the optical XCVRs **358** and data engine **362** in the IO drawer **106** over both of the low byte cable **302** and high byte cable **304**.

**[0056]** The wire engine **328** of local control FPGA **210** receives the state of certain bits of the registers **324** or DC voltage signals and utilizes the data engine **322** to transmit these states to the registers **364** of the remote control FPGA **250**. The registers **324** include a predefined bit to assert the state of the PCIE auxiliary PERST DC voltage signal output from a PHB **130** to a device attached to the respective PCIE link, and a bit to receive the state of PCIE auxiliary device present DC voltage signal from a PCIE slot **152** in the IO drawer **106** connected to the PCIE link over the low byte cable **302** and high byte cable **304**. When the state of certain bits of registers **324** changes, the wire engine **328** of the local FPGA **210** automatically communicates these to registers **364** of the remote FPGA **250**. The wire engine **368** of remote control FPGA **250** receives the state of certain bits of the registers **364** or DC voltage signals and utilizes the data engine **362** to transmit these states to the registers **324** of the local control FPGA **210**. Whenever the state of these certain bits of registers **364** changes, the wire engine **368** automatically communicates these to registers **324** of the local control FPGA **210**.

**[0057]** The respective wire engine **328**, **368** on each on each end of the optical cables **302**, **304** provide an alternative signaling mechanism for PCIE auxiliary signals or other DC voltage signals within the fiber optic cables to establish or receive the active or inactive state of the auxiliary signals at the respective other end of the cable.

**[0058]** The registers **324** of local control FPGA **210** include bits representing various properties of the cable card **140**, such as the type of the cable card itself, the type and connection states of the low byte cable **302** and high byte cable **304**. The registers **324** include bits to detect the states of certain hardware inputs from or control the states of certain hardware outputs to the components of the cable card **140**. The registers **324** of local control FPGA **210** include bits representing various properties of the cable connections to that IO drawer,

such as representing which location on the IO drawer **106** of the low byte cable **302** and high byte cable **304** are connected, to enable firmware to determine that cables are properly connected.

**[0059]** The cable card **140** and the IO drawer **106** optionally include Link Active LEDs **380** in association with each of the low byte cable **302** and high byte cable **304**. Firmware operating in the CEC **102** utilizes bits within the registers **324** of the local control FPGA **210** to active or deactivate the link active LEDs **380** to indicate that the cable is or is not actively transmitting signals between the cable card **140** and IO drawer **106**. Firmware operating in the CEC **102** performs other control and communications operations, such as activating or deactivating power to the IO drawer **106**, a bay **150**, or **154**, PCIE slots **152**, or other components within the IO drawer **106**.

**[0060]** Referring to FIG. **4**, there are shown example operational features generally designated by the reference character **400** of a physical layer view of the example system **100** of FIG. **1** in accordance with a preferred embodiment. An I2C multiplexer (MUX) **402** and a respective data engine **404**, **414** and an I2C SPI flash **410**, **420** at host side FPGA **210** and IO drawer side FPGA **250** are shown connected by CXP lanes. A respective SPI flash **412**, **422** is coupled to the I2C SPI flash logic **410**, **420** at host side FPGA **210** and IO drawer side FPGA **250**. The I2C multiplexer (MUX) **402** and the respective data engine **404**, **414** enable access secondary I2C buses in a consistent manner. The secondary I2C bus **254** could be in the IO drawer **106** talking to the drawer controller **252** or it could be the I2C bus **204** to each of the cable connector.

**[0061]** Referring to FIG. **5**, there is shown a flow chart illustrating example firmware operational features using sideband status information of the example system **100** in accordance with a preferred embodiment.

**[0062]** As indicated in a decision block **500**, checking for a PCIE card present in CEC PCIE slot is performed. When a PCIE card is not present in CEC PCIE slot, then operations return as indicated in a block **501**. When a PCIE card is present in CEC PCIE slot, the cable card present port expander is read as indicated in a block **502**. As indicated in a decision block **504**, again checking for the PCIE cable card present in CEC PCIE slot is performed. When a PCIE cable card is not present in CEC PCIE slot, then operations return at block **501**.

**[0063]** When a PCIE cable card is present in CEC PCIE slot, the local FPGA registers are read to get cable status and connection locations as indicated in a block **506**. Checking if the cable card includes a pair of cables or dual cables is performed as indicated in a decision block **508**. When the cable card includes dual cables, checking is performed to determine if the high byte control is working as indicated in a decision block **510**. If the high byte control is not working, a cable error is logged for service as indicated in a block **512**. If the high byte control is working and when the cable card does not include dual cables, checking is performed to determine if the low byte control is working as indicated in a decision block **514**. If the low byte control is not working, a cable error is logged for service as indicated in a block **516**.

**[0064]** As indicated in a block **518**, local FPGA registers are read to verify cables are correctly connected low to low, high to high, and same PCIE link connection at the IO drawer. Checking if cabled correctly is performed as indicated in a decision block **520**. If not cabled correctly, a cable error is logged for service as indicated in a block **522**. If cabled correctly, PERST to the IO drawer PCIE link connection is de-asserted as indicated in a block **524**. Checking is performed to determine if connected to PCIE switch in IO drawer as indicated in a decision block **526**. When connected to PCIE switch in IO drawer, the PCIE switch and downstream PCIE links to PCIE slots in the IO drawer are configured as indicated in a block **528**. The PCIE slot or PCIE slots in the IO drawer under the PHB connected to this cable are configured as indicated in a block **530** following block **528** and following decision block **526** if not connected to PCIE switch in IO drawer. The PCIE link connection configuration is complete as indicated in a block **532**.

**[0065]** Referring to FIGS. **6**A and **6**B, there is shown example PCI bus generally designated by the reference character **600** of the example system **100** in accordance with a preferred embodiment. The illustrated PCI bus **600** provides a path from a PCIE host bridge (PHB) **602** to an optical cable attached IO drawer **106**. For example, two 12 lane optical cables are used to carry PCIE and sideband traffic between the system PHB **602** and the IO drawer **106**. Each cable contains 8 lanes of PCIE traffic, 1 channel of sideband communication and 1 spare optical channel with two channels unused. With one cable carrying PCIE lanes 0-7 (low byte) and the other carrying PCI lanes 8-15 (high byte) together they make a x16 PCIE bus connection to the IO drawer **106**. The one spare optical channel per cable in each direction allows firmware to route a single PCIE lane away from a faulted channel to the spare channel when firmware determines a channel that is faulted. Other implementations could use, for example two unused channels as spares for the PCIE links as well.

**[0066]** As shown in FIG. **6**A, PCI bus **600** includes a PCIE retimer **604**, field programmable gate array (FPGA) communication devices **606**, lane sparing multiplexers (muxes) **608** including lane sparing transmit (TX) muxes **610**, and lane sparing receive (RX) muxes **612**, and High and Low optical module transceivers (CXPs) **614**, **616** with respective optical connections X9+DN, SB, X9+UP, SB. For example, with each optical cable containing 12 channels of bidirectional traffic, eight of those lanes optionally are used for PCIE, one used for sideband communication SB and one of the unused channels wired to be used as a spare PCIE lane. As shown, the FPGA communication devices **606** are connected to the PCIE retimer **604**, High and Low optical module transceivers (CXPs) **614**, **616**, and the lane sparing multiplexers (muxes) **608**.

**[0067]** As shown in FIG. **6**B, PCI bus **600** includes High and Low optical module transceivers (CXPs) **620**, **622** that are connected to the High and Low optical module transceivers (CXPs) **614**, **616** of FIG. **6**A with the respective optical connections X9+DN, SB, X9+UP, SB. PCI bus **600** includes lane sparing multiplexers (muxes) **624** including lane sparing transmit (TX) muxes **626**, and lane sparing receive (RX) muxes **628**; a PCIE retimer **630**, field programmable gate array (FPGA) communication devices **632**, and a PCIE switch **634** of the IO drawer **106**. As shown, the PCIE switch **634** is connected to a plurality of PCI slots **636** #1-N, for example, 6 PCI slots **636** of the IO drawer **106**. As shown, the FPGA communication devices **632** are connected to the PCIE retimer **630**, High and Low optical module transceivers (CXPs) **620**, **622**, lane sparing multiplexers (muxes) **624**, and connected to the PCIE switch **634**, a drawer controller (CMC) **638** and the PCI slots **636** of the IO drawer **106**.

**[0068]** Referring to FIG. **7** there is shown example lane sparing hardware generally designated by the reference char-

acter **700** of the example system **100** in accordance with a preferred embodiment. Lane sparing hardware **700** is used to reroute PCIE lanes to a spare optical channel on both ends of the optical path in accordance with a preferred embodiment. As shown, lane sparing hardware **700** includes an I2C port expander device **702** coupled to an optical transceiver (CXP) connector **704** by a CXP interface block **706**.

**[0069]** The I2C port expander device **702** is used to control a set of PCIE lane multiplexers (muxes) **708**, **710**. As shown, lane sparing hardware **700** includes a retimer **714** coupled to the PCIE lane muxes **708**, **710**, a field programmable gate array (FPGA) **716** coupled by sideband and SM bus to the CXP interface block **706** and CXP connector **704** and coupled to the I2C port expander device **702**. Lane sparing hardware **700** includes control and status registers **718** included with the CXP connector **704**.

**[0070]** Interrogating PCIE retimers, such as retimer **714** can be used to determine the faulted optical channel. For example, by reading retimer registers it is possible to determine for each PCI lane what stage of the PCIE link training sequence has been initiated. This information can be used to determine if one lane has not progressed at all which indicates a bad optical channel. By reading transmit status registers in the optical transceivers it is possible to determine for each if a fault has occurred. This information can be used to determine a bad optical channel. By reading receiver status registers **718** in the optical transceivers **704** it is possible to detect a loss of signal from the host and this information can be used to determine a bad optical channel.

**[0071]** By programming muxes in the PCIe path firmware can reroute PCIE away from the faulted channel. The I2C port expander device **702** is controlled through the FPGA **716**, for example using I2C communication through the sideband signals. When the I2C port expander device **702** is programmed by firmware it changes the input to output mapping of the PCIE lane muxes **708**, **710** which causes the PCIe bus to be rerouted down a different optical channel. When the operation is complete a reset sequence can be initiated on the PCIE bus to recover the link to its optimal width.

**[0072]** As illustrated and described with respect to FIGS. **8**A, **8**B, and **8**C, the health check includes verifying the current link speed and width and comparing to the reported speed and width. The health check includes verifying status of a local field programmable gate array (FPGA) including the sideband signals. The health check includes verifying status of a remote field programmable gate array (FPGA). The health check includes verifying status of a Host Bridge Link (HBL). The health check includes verifying status of optical transceivers, and active optical cable (AOC). The health check includes verifying status of PCI retimers.

**[0073]** In accordance with features of the invention, the health check is restarted after a specified time period. For example, a process is started that initiates the health check at one minute intervals. The PCIE health check is run during the initial IPL of the system ensuring the links are running at their maximum capability. The PCIE health check is run during the recovery of errors related to the PCI link. The health check is executed during the recovery of PCIE correctable errors, PCIE uncorrectable errors and PHB link related errors including link down. If the health check discovers a correctable active optical cable (AOC) error the link will be repaired and the connection reset. The AOC error takes precedence over an original reported error.

**[0074]** Referring to FIGS. **8**A, **8**B, and **8**C, there is shown a flow chart illustrating example operations for implementing heath check for optical cable attached PCIE enclosures in accordance with a preferred embodiment starting at a block **800** in FIG. **8**A.

**[0075]** As indicated in a block **802**, optical and retimer components are interrogated to determine if there was a PCIE reset or if there is degradation of the PCIE link or if there was an initial program load (IPL). As indicated in a decision block **804** checking if there was a PCIE link reset from an initial program load (IPL) is performed. When the PCIE link was not reset from an initial program load (IPL) as indicated in a decision block **806** checking if there was a PCIE link reset from error recovery or as a result of a concurrent maintenance operation is performed. If not, as indicated in a decision block **808** checking if there was a PCIE link reset from PCIE link degradation is performed. If not, then the operations return as indicated in a block **812**. When the PCIE link was reset at either of decision blocks **804**, **806** or **808**, then a control signal, such as PERST is used to reset the PCIE module or PCIE enclosure, then waiting for a settling delay to perform PCIE link health check as indicated in a block **810** where the link width and speed will be verified and retrained if this link is not optimal.

**[0076]** A health check includes verifying the current link speed and width as indicated in a block **814** and comparing to the reported speed and width to determine if the current PCIE link speed and width are within acceptable parameters as indicated in a decision block **816**. When current PCIE link speed and width are within acceptable parameters, then as indicated in a block **818** operations go to block **822** in FIG. **8**B. When current PCIE link speed and width are not within acceptable parameters, then as indicated in a block **820** operations go to block **852** in FIG. **8**C.

**[0077]** Following block **822** in FIG. **8**B, health check continues with verifying the status of the local FPGA including the sideband signals as indicated in a block **824**. Checking if the status of the local FPGA including the sideband signals is acceptable is performed as indicated in a decision block **826**. When status of the local FPGA is not acceptable, then as indicated in a block **841** operations go to block **852** in FIG. **8**C.

**[0078]** When status of the local FPGA is acceptable, then verifying the status of the remote FPGA is performed as indicated in a block **828**. Checking if the status of the remote FPGA is acceptable is performed as indicated in a decision block **830**. When status of the remote FPGA is acceptable, then verifying the status of the Host Bridge Link is performed as indicated in a block **832**. Checking if the status of the Host Bridge Link is within acceptable parameters is performed as indicated in a decision block **834**. Otherwise as indicated at block **841** operations go to block **852** in FIG. **8**C. When the status of the Host Bridge Link is within acceptable parameters, then verifying the status of the optical transceivers (CXP) and active optical cable (AOC) are performed as indicated in a block **836**. Then as indicated in a block **840** operations go to block **842** in FIG. **8**C. Otherwise as indicated in a block **841** operations go to block **852** in FIG. **8**C.

**[0079]** Following block **842** in FIG. **8**C, checking if the status of the optical transceivers (CXP) and active optical cable (AOC) are within acceptable parameters is performed as indicated in a decision block **844**. When the optical transceivers (CXP) and active optical cable (AOC) are within acceptable parameters, then verifying the status of the PCI

retimers is performed as indicated in a block **846**. Checking if the status of the PCI retimers is within acceptable parameters is performed as indicated in a decision block **848**. If so, then as indicated in a block **850** the health check is restarted after a specified time period, such as after 1 minute.

**[0080]** Following block **852** in FIG. **8**C, and when the optical transceivers (CXP) and active optical cable (AOC) or the PCI retimers are not within acceptable parameters, then PCIE link recovery is performed and any correctable errors are corrected as indicated in a block **854**. The PCIE link is repaired and the connection reset as indicated in a block **856**. Checking is performed to determine if the PCIE link is operable as indicated in a decision block **858**. If the PCIE link is operable then as indicated at block **850** the health check is restarted after a specified time period. Otherwise if the PCIE link is not operable, an error code is generated to provide a single failure to the customer as indicated in a block **860**.

**[0081]** For example, an Error Identification (ID) related to the health check is correlated with Error ID generated from error handling code for the Root Complex PCIE Host Bridge (PHB) or the first upstream device in the enclosure to provide a single failure to the customer. When the health check is initiated by an error recovery mechanism for the PHB or the PCIE Switch, and the health check finds an issue in the link path, the errors that are reported are all properly correlated with each other, resulting in one error ID in the logs. Error IDs will be propagated from the error recovery handlers to the health check, to ensure these errors stay correlated.

**[0082]** Referring now to FIG. **9**, an article of manufacture or a computer program product **900** of the invention is illustrated. The computer program product **900** is tangibly embodied on a non-transitory computer readable storage medium that includes a recording medium **902**, such as, a floppy disk, a high capacity read only memory in the form of an optically read compact disk or CD-ROM, a tape, or another similar computer program product. Recording medium **902** stores program means **904**, **906**, **908**, and **910** on the medium **902** for carrying out the methods for implementing health check of preferred embodiments in the system **100** of FIG. **1**.

**[0083]** A sequence of program instructions or a logical assembly of one or more interrelated modules defined by the recorded program means **904**, **906**, **908**, and **910**, direct the computer system **100** for implementing health check of optical cable attached PCIE enclosures or IO drawer **106** of a preferred embodiment.

**[0084]** While the present invention has been described with reference to the details of the embodiments of the invention shown in the drawing, these details are not intended to limit the scope of the invention as claimed in the appended claims.

**1-11**. (canceled)

**12**. A computer system for implementing health check for optical cable attached Peripheral Component Interconnect Express (PCIE) enclosures, comprising:

a processor;

a PCI host bridge (PHB); said PHB connected to said processor and one or more optical cables connecting between the PHB and an PCIE enclosure providing a PCIE link to the PCIE enclosure; and system firmware for implementing health check functions;

said processor using said system firmware for performing the steps of:

resetting a PCIE link to the PCIE enclosure responsive to identifying a predefined event;

responsive to resetting the PCIE link, providing a set delay and performing a link health check for verifying PCIE link width and speed.

**13**. The system as recited in claim **12**, includes control code stored on a computer readable medium, wherein said control code comprising said system firmware.

**14**. The system as recited in claim **12**, wherein resetting the PCIE link to the PCIE enclosure responsive to identifying a predefined event includes identifying an initial program load (IPL) of the system.

**15**. The system as recited in claim **12**, wherein resetting the PCIE link to the PCIE enclosure responsive to predefined event includes identifying a recovery from an error or identifying a concurrent maintenance operation.

**16**. The system as recited in claim **12**, wherein resetting the PCIE link to the PCIE enclosure responsive to predefined event includes a process initiating the health check at a specified time interval.

**17**. The system as recited in claim **12**, wherein resetting the PCIE link to the PCIE enclosure responsive to predefined event includes identifying link degradation.

**18**. The system as recited in claim **12**, performing the link health check verifying PCIE link width and speed includes at least one of verifying status of a local field programmable gate array (FPGA) including the sideband signals, verifying status of a remote field programmable gate array (FPGA), verifying status of a Host Bridge Link (HBL) including verifying status of optical transceivers, and active optical cable (AOC), and verifying status of a PCI retimers.

**19**. The system as recited in claim **12**, wherein performing the link health check for verifying PCIE link width and speed includes performing the link health check during the recovery of PCIE correctable errors, PCIE uncorrectable errors and recovery from PHB link errors including link down.

**20**. The system as recited in claim **12**, wherein performing the link health check for verifying PCIE link width and speed includes performing the link health check during an initial program load (IPL) of the system.

* * * * *